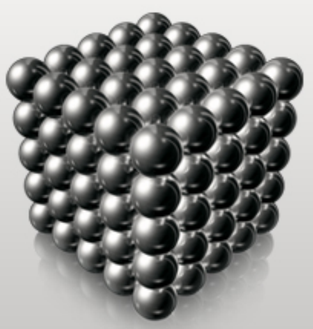
HPC Technology Update

Session 609

These are confidential sessions—please refrain from filming or taking pictures.



Enterprise IT

HPC Technology Update

Skip Levens
Server Technology Manager

http://developer.apple.com/wwdc2005/

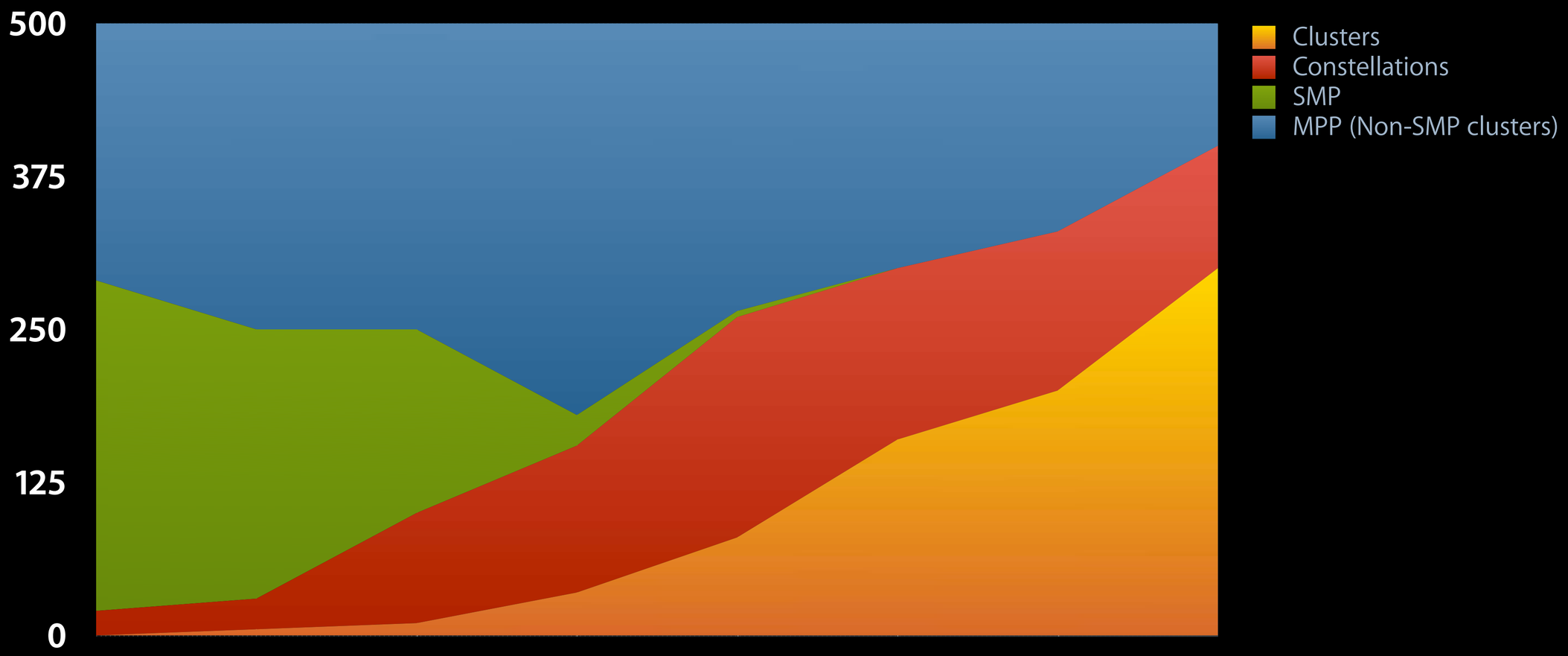
What We'll Cover

- HPC trends and Apple deployments
- HPC technology survey, including:
  - Interconnects—Myrinet, SilverStorm
  - Tools—Absoft Update
- Technology demo
- Wrapup

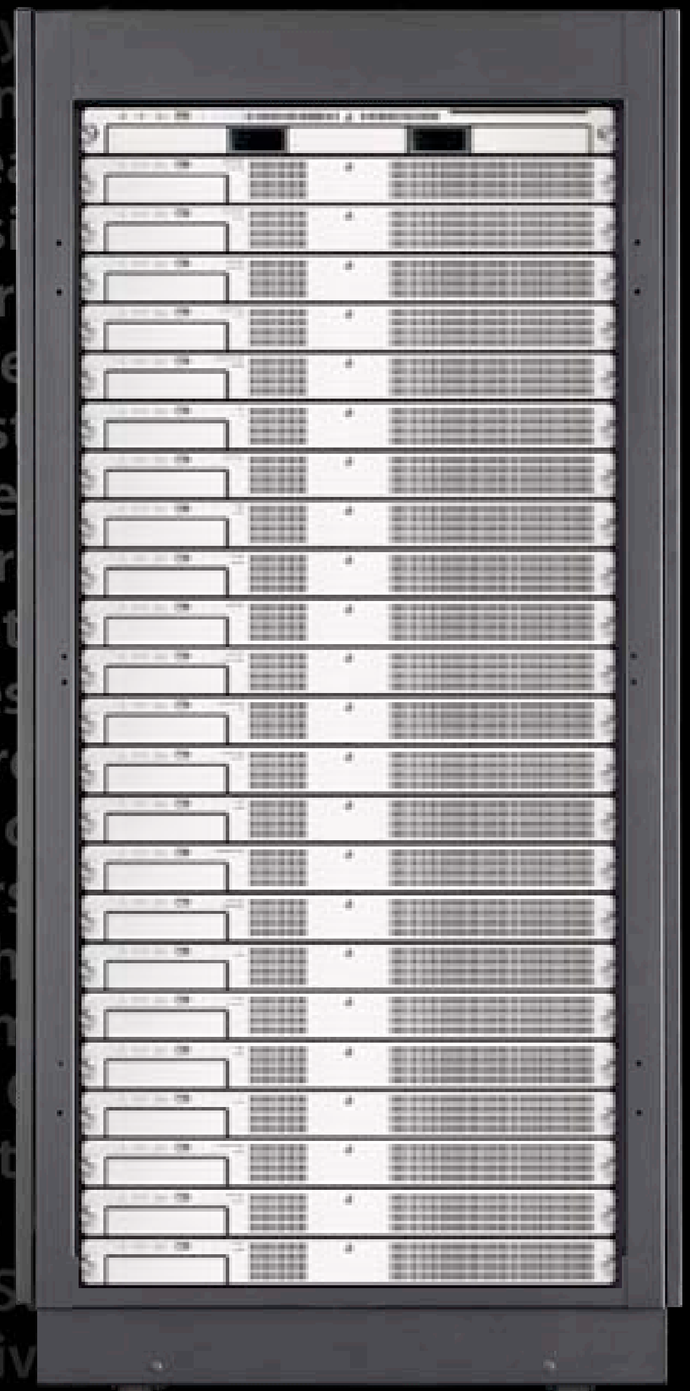
HPC Trends and Apple Deployments

http://developer.apple.com/wwdc2005/

Supercomputer Architectures Since 1998



Virginia Tech University • University of Illinois • University of California Los Angeles • Yale University • Princeton University • University • Danforth Plant Institute • Sandia National Labs • Mayo Clin California Davis • Texas A&M University • Naval Medical Resea • Pacific Title • University of Michigan • Georgetown Universi Virginia • University of California San Diego • Dana Farber • Purdue University • Beth Israel Medical Center • McGill Unive of North Carolina • Temasek Life Sciences Laboratory • Aust Defence • Jet Propulsion Laboratory • Outpost Digital • Unive University of Colorado • University of Albany • Mitre Corpor Livermore National Labs • Beth Israel Deaconess Medical Cent Pittsburgh • Academia Sinica • Monterey Bay Aquarium Res Purdue University • United States Department of Agricultur Reserve University • Harvard University • National Institute o Health Sciences • Idaho State University • Stanford Univers University • Children's Research Institute • University of Mich of Idaho • Queens College • Hunter College • Bingham University of Oregon • Rockefeller University • Hamilton C University • University of North Carolina Charlotte • Universit Madison • Furman University • University of Kentucky • University of Toronto • University of Iowa • Washington S University of Florida • University of Michigan Dearborn • Univ Washington State University • University of Southern California • Colorado University • University of Maine • University of Chicago • COLSA Corporation •



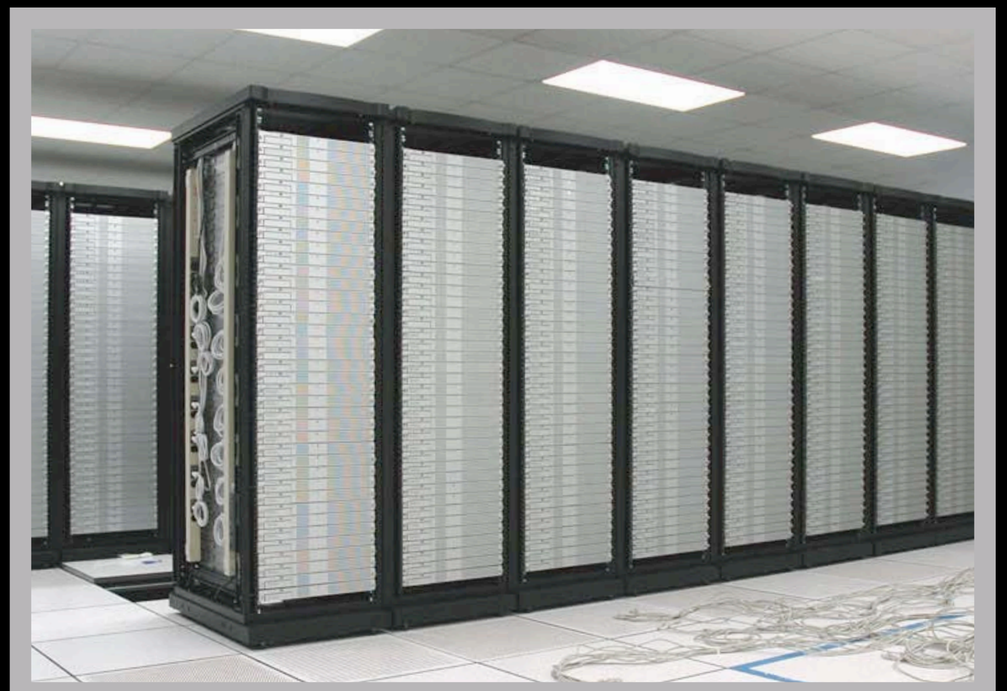
Virginia Tech—'System X'

- Converted from Dual 2.0 GHz PowerMac workstations to 2.3 GHz Xserve G5's
- Footprint reduced by 2/3
- 425KW load reduced to 310KW
- Generated 12.25 Teraflop Run to qualify at #7 on Top500, remains #1 Academic Cluster
- Entered production and helping drive new applications to Mac OS X

UCLA—Dawson Cluster

- 256 Xserve G5 nodes (128 2.0 GHz and 128 2.3 GHz)
- GigE and LAM/MPI
- Entered production quickly
- Generated .948 Teraflop run to qualify as #444 on Top500 list
- Have since generated a 1.21 TF run using Dauger Research PoochPro and MacMPI

COLSA—MACH 5



- 1562 Dual 2.3 GHz Xserve G5 Cluster Nodes
- Gigabit Ethernet
- Hypersonic Missile Research
- Highly dense CFD codes

UIUC—Turing Cluster

- Used by Computational Science and Engineering and Center for Simulation of Advanced Rockets
- 640 nodes (plus 10 utility servers)
- Myrinet interconnect
- $3M for 10 Teraflop peak (includes nodes, network and data center cooling and power upgrades)

Bowie State—XSEED

- 224 Dual 2.0 Xserve G5 Nodes
- Myrinet Interconnect
- 2.104 Teraflop benchmark run
- Student and faculty planned and deployed

Apple Workgroup Cluster

- Turnkey Cluster Computing
- Within reach of workgroup, department or discretionary budget
- Scale to fit your needs

HPC Technology Survey

http://developer.apple.com/wwdc2005/

Applications

Tools

Resource Management

File Systems

Code Libraries

IPC Middleware

Hardware Interconnects

OS Platform

Hardware Platform

Applications

Tools

Resource Management

File Systems

Code Libraries

IPC Middleware

Hardware Interconnects

OS Platform

Hardware Platform

Hardware Platform

Xserve G5 and Xserve G5 Cluster Node

- Basic building block
- 64-Bit computing and memory
- Capable of up to 35 Billion single-precision or 10 Billion floating point instructions per second
- Well-balanced design of I/O throughput, environmentals and processing power
- Now supports 2GB DIMMS for up to 16GB of ECC memory

Applications

Tools

Resource Management

File Systems

Code Libraries

IPC Middleware

Hardware Interconnects

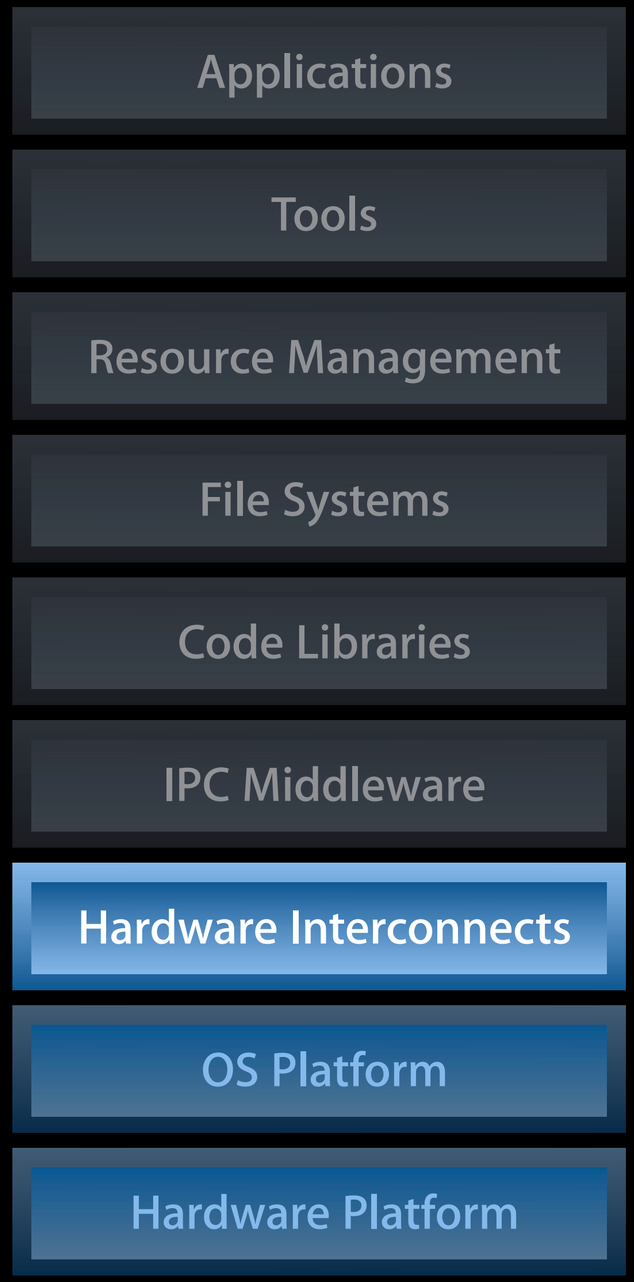
OS Platform

Hardware Platform

Tiger Server

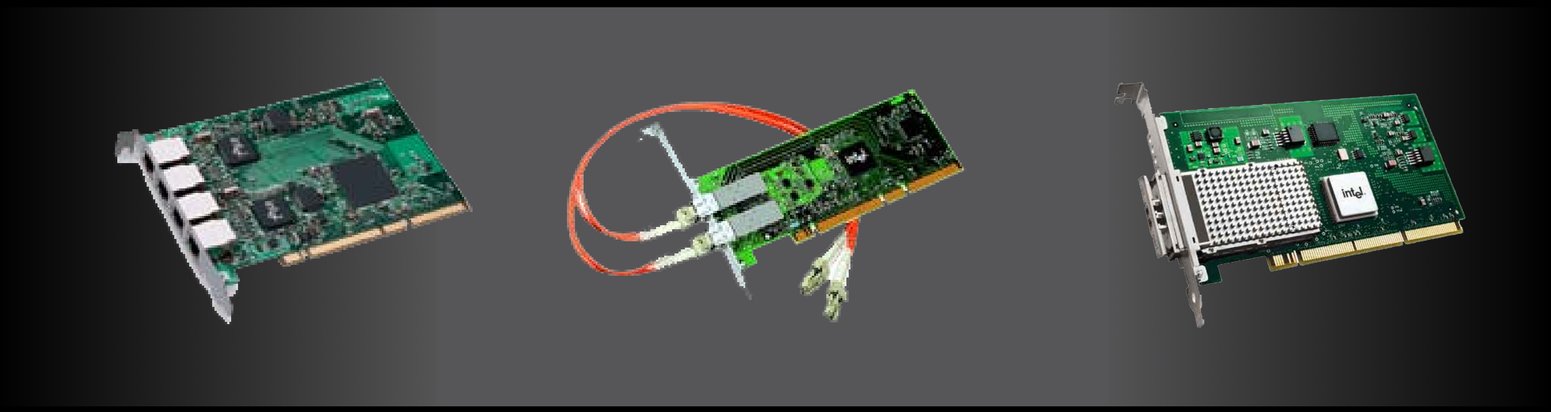
HPC Benefits

- 64-bit memory address space
- Xcode 2
  - gcc 4.0
- Kernel tuning via launchd now
- Memory Manager Code

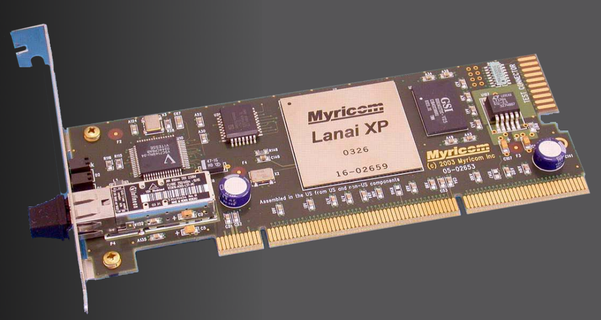

Applications
Tools
Resource Management
File Systems
Code Libraries
IPC Middleware
Hardware Interconnects
OS Platform
Hardware Platform

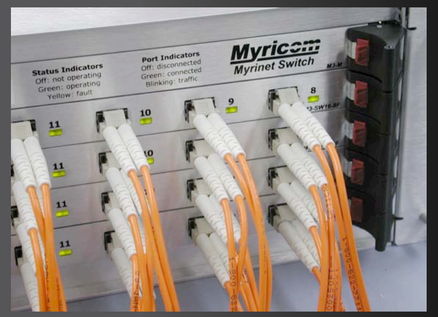
Hardware Interconnects—Gigabit Ethernet

- GigE
- Multiport GigE Expansion Cards, optical GigE, 10GigE



Hardware Interconnects—Myrinet





Myrinet

What's new?

- Two new Xserve clusters with Myrinet will appear in the June 2005 TOP500 Supercomputer list
- Myrinet Express (MX)
  - A new message-passing system for Myrinet clusters
  - MX-2G is released for Myrinet PCI-X NICs
  - Much lower latency than GM
    - 3.6µs raw MX, 3.8µs MPICH-MX, measured on G5s with E or F cards
  - Great ethernet-emulation performance
  - Complete set of middleware
    - MPICH2-MX coming soon

Myrinet Express (MX) and Mac OS X

- MX uses a new approach to memory management
  - Small and medium messages (<32KB) are copied
  - Large messages are pinned/unpinned on demand
  - Tiger's improved U3 DART code makes this practical on Mac OS X
  - 4x as much memory is available for communication vs GM
    - 375MB with GM vs 1.5GB on MX
  - Linking complex MPI applications is as easy with MPICH-MX as it is with a sockets-based MPI because there is no need to override malloc/free
- Tested with Apple's gcc-4
- Experimental support for 64-bit binaries on Tiger
- The ethernet-emulation interface with Tiger can be configured via System Preferences, just like any other ethernet interface

Hardware Interconnects—InfiniBand

- SilverStorm Technologies and Small Tree Communications

SilverStorm Backgrounder

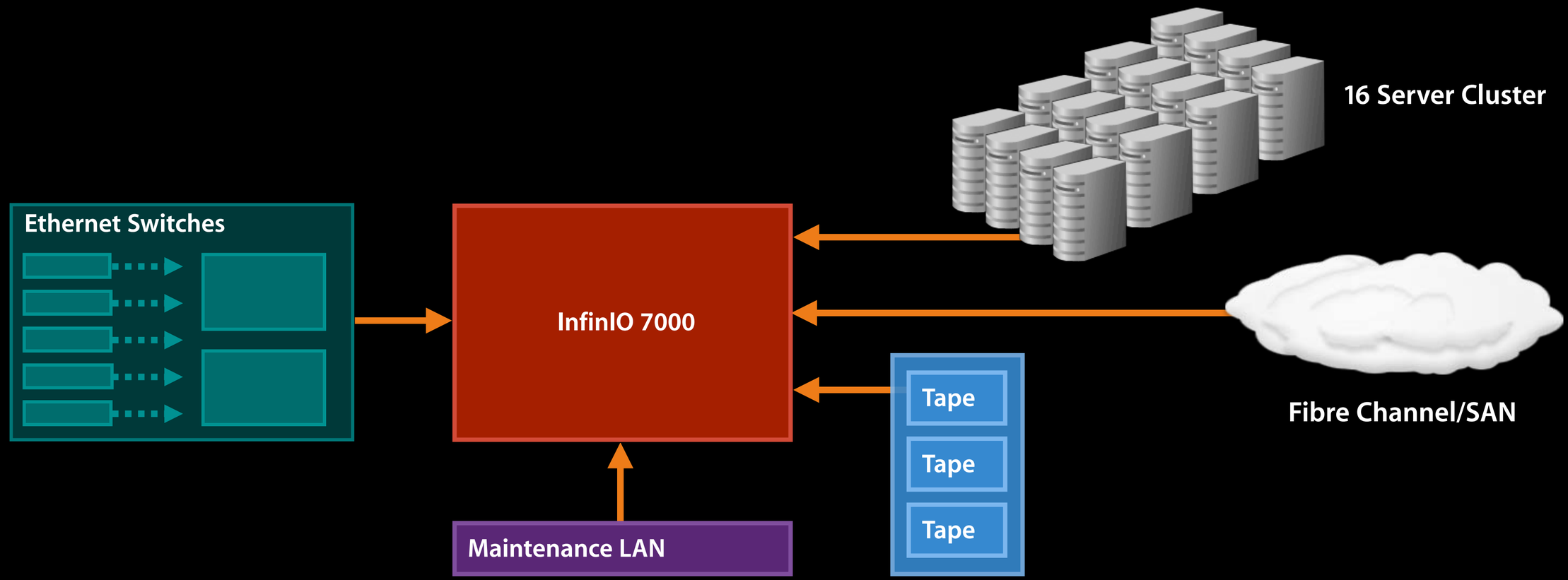
- Founded in June 2000 (formerly InfiniCon Systems)
- Sales and support teams across the US, Europe, and Asia
- Proven management and development experience in delivering enterprise-class networking and server I/O products
- Complete end-to-end solutions for HPC and enterprise computing
  - Shared I/O Systems—InfiniBand, Fibre Channel, Ethernet
- Partnered with Apple and Small Tree Communications to deliver a complete software solution for Mac OS X
- Well funded with over $50M raised

Small Tree Backgrounder

- Hardware and Software development experience from Cray, SGI, Sun, and HP
- Technology Partnerships include Intel, Silicom, Apple, and SilverStorm
- System Integration expertise for HPC
- Wide range of offerings including GigE, multiport GigE, 10GigE, and InfiniBand
- Focused exclusively on Mac OS X platform

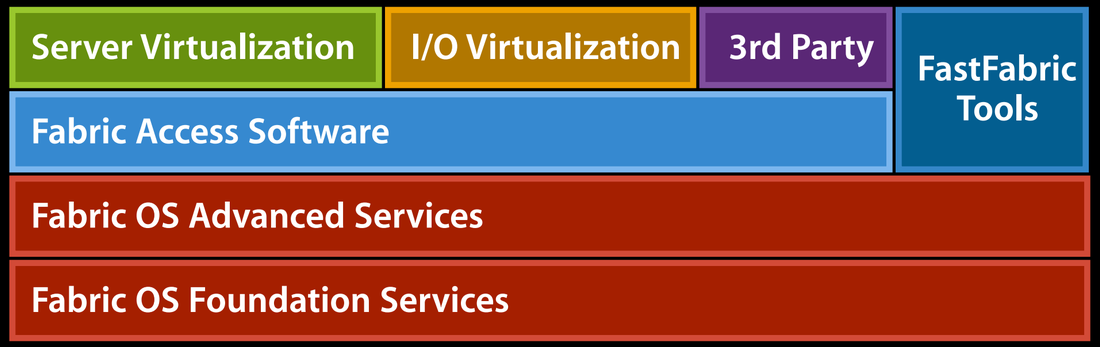
Vision for High Performance Computing

- ‘One Wire’ Data Center Computing
- High Performance Commercial Computing
- High Performance Technical Computing



SilverStorm Technologies Product Family

End-to-end networking solutions



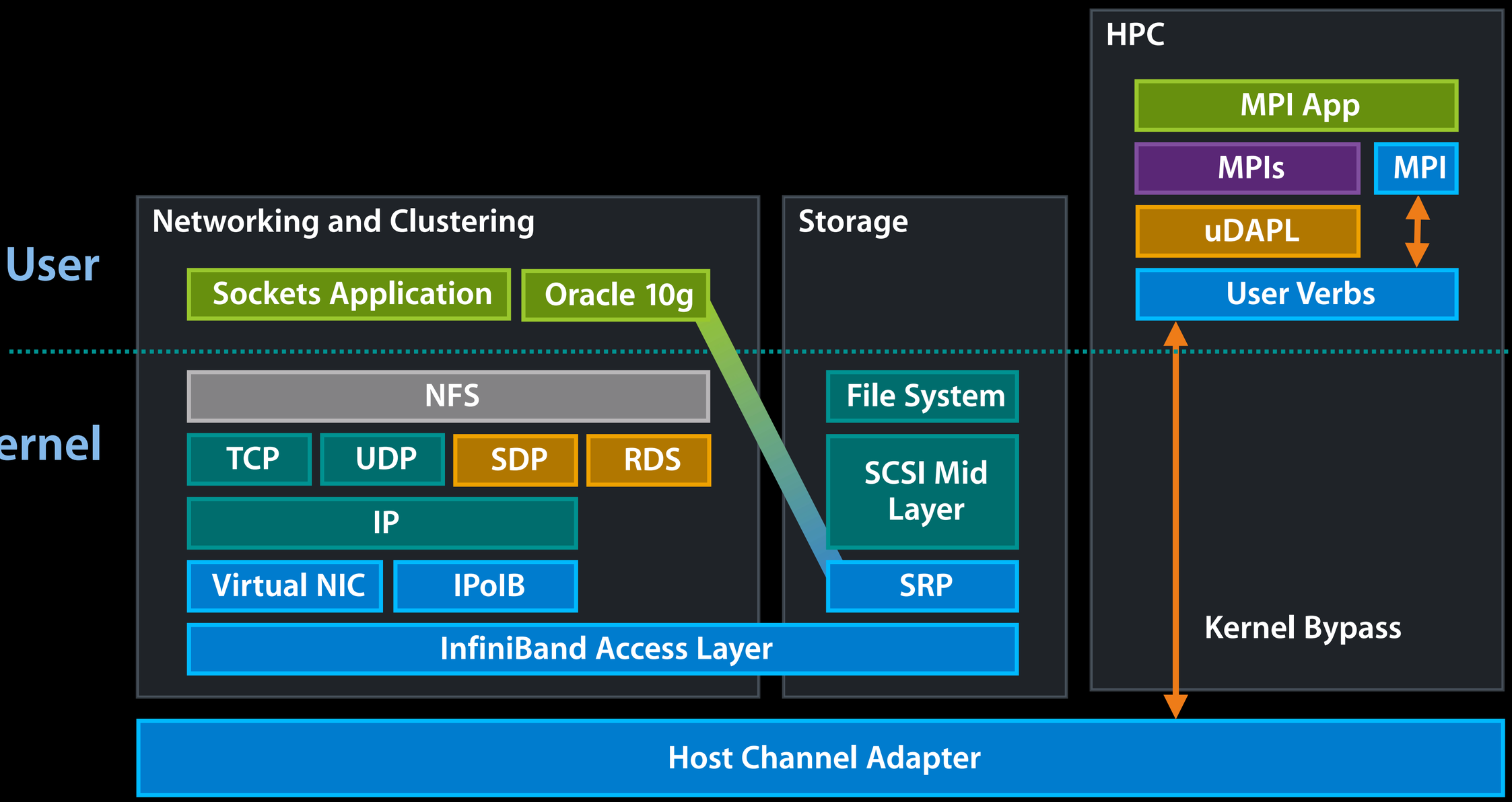
QuickSilver SW Architecture

Flexible Switch Configurations

InfiniBand HCAs

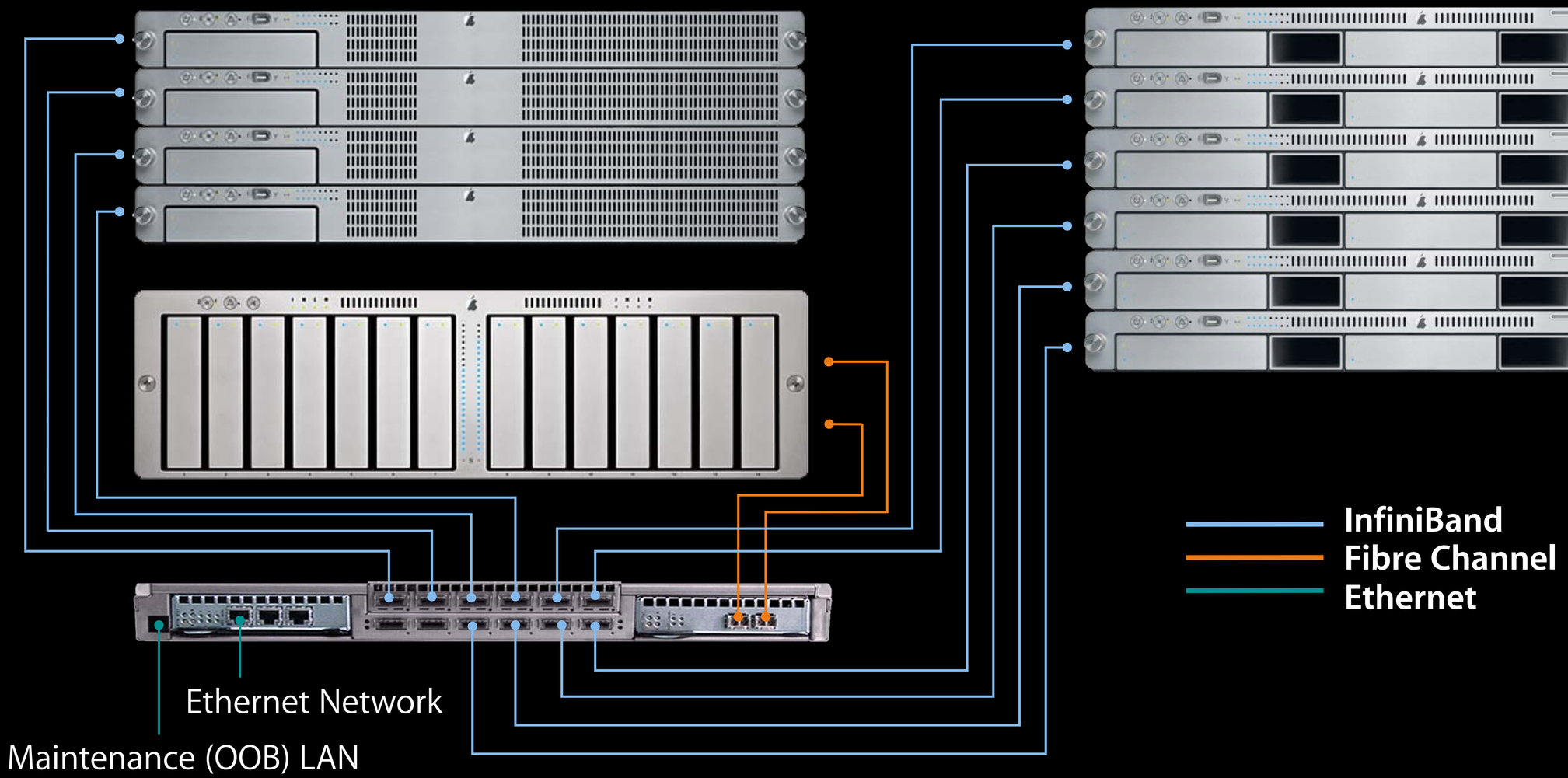
Virtual I/O

Highly Scalable Host Software Infrastructure



High Performance Network Consolidation

InfiniBand as single server connection for IPC, storage and networking



Strategic Software and Technology Partnerships

MPI

- OHIO STATE
- Verari systems
- SCALI High Performance Clustering
- intel

Manufacturing

- LSTC Livermore Software Technology Corp.
- CD adapco Group
- ABAQUS
- ESI GROUP THE VIRTUAL TRY-OUT SPACE® COMPANY
- FLUENT
- ANSYS

Oil & Gas

- WesternGeco
- Landmark A Halliburton Company
- VERITAS
- Schlumberger

Life Sciences

- BLAST
- Gaussian.com THE OFFICIAL GAUSSIAN 03 WEB SITE

Data Base

- ORACLE

Applications

Tools

Resource Management

File Systems

Code Libraries

IPC Middleware

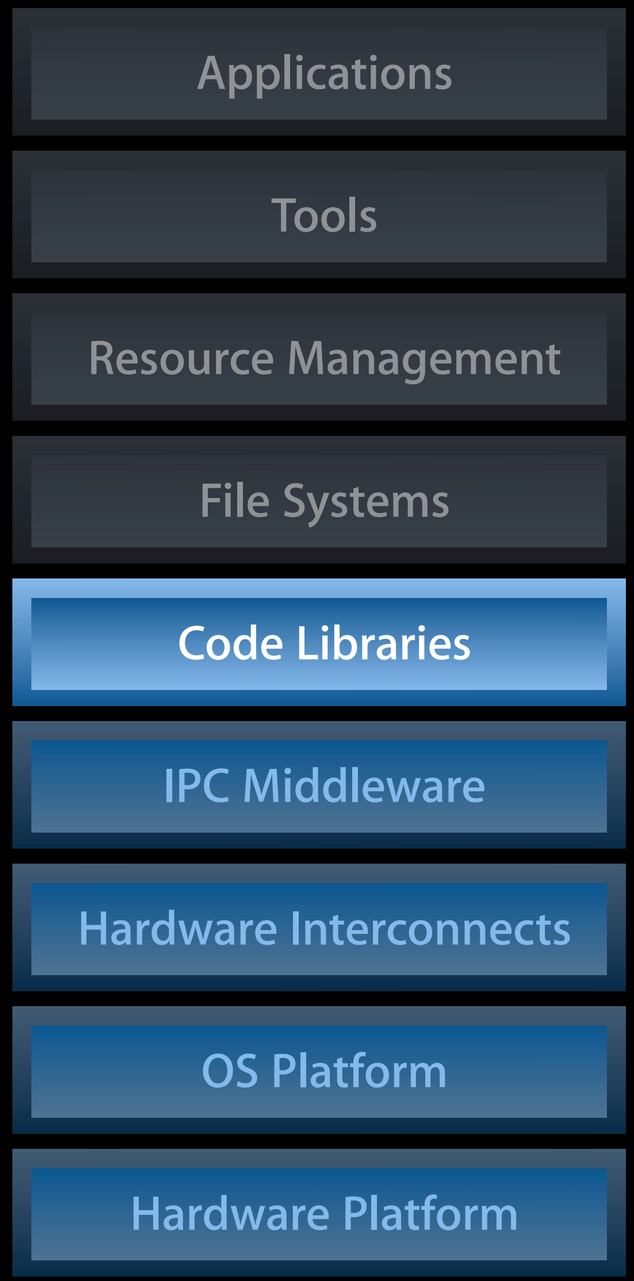
Hardware Interconnects

OS Platform

Hardware Platform

IPC Middleware

- LAM/MPI
- Open MPI
- MPICH from Argonne National Labs
- MVAPICH from Ohio State
- MPI/PRO from MPI Software Technology
- MacMPI_X and LnxMPI_S from UCLA
- Globus
- Linda/Paradise from Scientific Computing Associates
- Massively Parallel
- SilverStorm/Small Tree MPI
- PVM from Oak Ridge National Laboratory
- XLR8 from GridIron Software
- MPICH GM/MPICH MX


Applications
Tools
Resource Management
File Systems
Code Libraries
IPC Middleware
Hardware Interconnects
OS Platform
Hardware Platform

Code Libraries

- Accelerate Framework from Apple
  - BLAS Levels 1, 2, 3 (ATLAS tuned, selectively SMP aware)
  - LAPACK, linear systems and eigenvalue problems
  - Digital Signal Processing, 1D, 2D FFT's [vDSP]
- Absoft IMSL Math Libraries
- NAG Math Libraries
- Goto BLAS
- Verari
- CCSM
- CTH

Applications

Tools

Resource Management

File Systems

Code Libraries

IPC Middleware

Hardware Interconnects

OS Platform

Hardware Platform

File Systems

- NFS
- AFP
- Xsan
- Lustre
- Blaze

Applications

Tools

Resource Management

File Systems

Code Libraries

IPC Middleware

Hardware Interconnects

OS Platform

Hardware Platform

Resource Management

- Schedulers
  - Xgrid
  - Platform LSF
  - PBS Pro from Altair Engineering
  - N1 Grid Engine from Sun
  - Maui/MOAB
  - Torque
  - Pooch
- Management and Monitoring
  - Ganglia
  - Big Brother from Quest Software
  - Oscar HA

Applications

Tools

Resource Management

File Systems

Code Libraries

IPC Middleware

Hardware Interconnects

OS Platform

Hardware Platform

Tools

Compilers/Debuggers/Profilers, Optimizers

- Xcode
  - gcc/gdb 4.0
- Absoft xlc/xlf
- Absoft Pro Fortran
- Absoft FxP
- Etnus TotalView

Absoft HPC Tools

Rodney Mach
Absoft HPC Technical Director
rwm@absoft.com

http://developer.apple.com/wwdc2005/

About Absoft

- Continuous tools support for Apple platforms since 1984
- First commercial Fortran and Fortran 95 compilers available on the Apple platform
- Continuing development of cutting edge developer and high performance tools for OS X Tiger

Absoft Solutions

- HPC SDK Basic Edition for Xserve clusters
- FxP MPI Debugger for OS X Tiger/Panther
- 64-bit and 32-bit Pro Fortran compiler for Tiger
- 64-bit and 32-bit IMSL Fortran Numerical Library v5.0 with serial and parallel support

HPC SDK for Xserve Clusters

- Includes high performance commercial compilers, debuggers, math libraries, MPI and other developer tools optimized for Xserve clusters
- Start developing HPC applications within seconds of installation
- Increase productivity with the seamless integrated HPC development environment

Best of Breed Components

- 64-bit and 32-bit Absoft Pro Fortran
- 64-bit and 32-bit IMSL Fortran Numerical Library v5.0 with serial and parallel support
- Absoft FxP MPI Debugging Solution
- MPICH2 and LAM/MPI
- VAST (Altivec and Automatic Parallelization)

FxP MPI Debugger Benefits

- Supports all major compiler vendors and MPI implementations in one package
- Easy to use graphical user interface is simple to learn and highly intuitive for immediate productivity and reduced training costs
- Builds on market leading Fx2 serial debugger for Fortran, C/C++, and assembler

FxP Features

- Basic MPI debugging support
- Automatically attach to MPI processes
- Connect to remote batch jobs for debugging production jobs
- View local and global variables, stack trace, registers, message queues, across all MPI processes
- Visually see the state of all running processes

Absoft Pro Fortran 9.2

- Supports 64-bit and 32-bit Fortran 95
- Free technical support
- Includes C/C++ compilers, IDE, Fx2 graphical debugger at no extra charge
- Fully link compatible with GNU compiler
- Includes BLAS, LAPACK90, ATLAS, HDF, Vax/Unix, and 2D/3D libraries

IMSL Fortran Numerical Library v5.0

- Complete collection of well documented mathematical and statistical algorithms for HPC, including 64-bit and 32-bit libraries; parallel and serial support
- MPI Parallel Processing support allows you to take advantage of your Xserve cluster
- MPI Complexity hidden from the user

Roadmap

- Continue to improve speed of our compilers, tools, and utilities
- Full Fortran 2003 support
- Introduction of new HPC tools
- Continued emphasis on our outstanding technical support
- Support for Intel-based Macs

Questions?

- For telephone inquiries, call: Jeff Livesay (248) 853-0050

Applications
Tools
Resource Management
File Systems
Code Libraries
IPC Middleware
Hardware Interconnects
OS Platform
Hardware Platform

Massively Parallel Technologies

Kevin Howard
President, CTO
Massively Parallel

http://developer.apple.com/wwdc2005/

Why Massively Parallel Technologies?

Massively Parallel Technologies (MPT) was founded to tackle the limiting factors in today's supercomputers

- MPT is based upon a new derivative of Amdahl's law that allows for superlinear performance
- MPT's proprietary communication models take full advantage of these new capabilities
- MPT is currently demonstrating linear and superlinear performance on Mac OS X Tiger

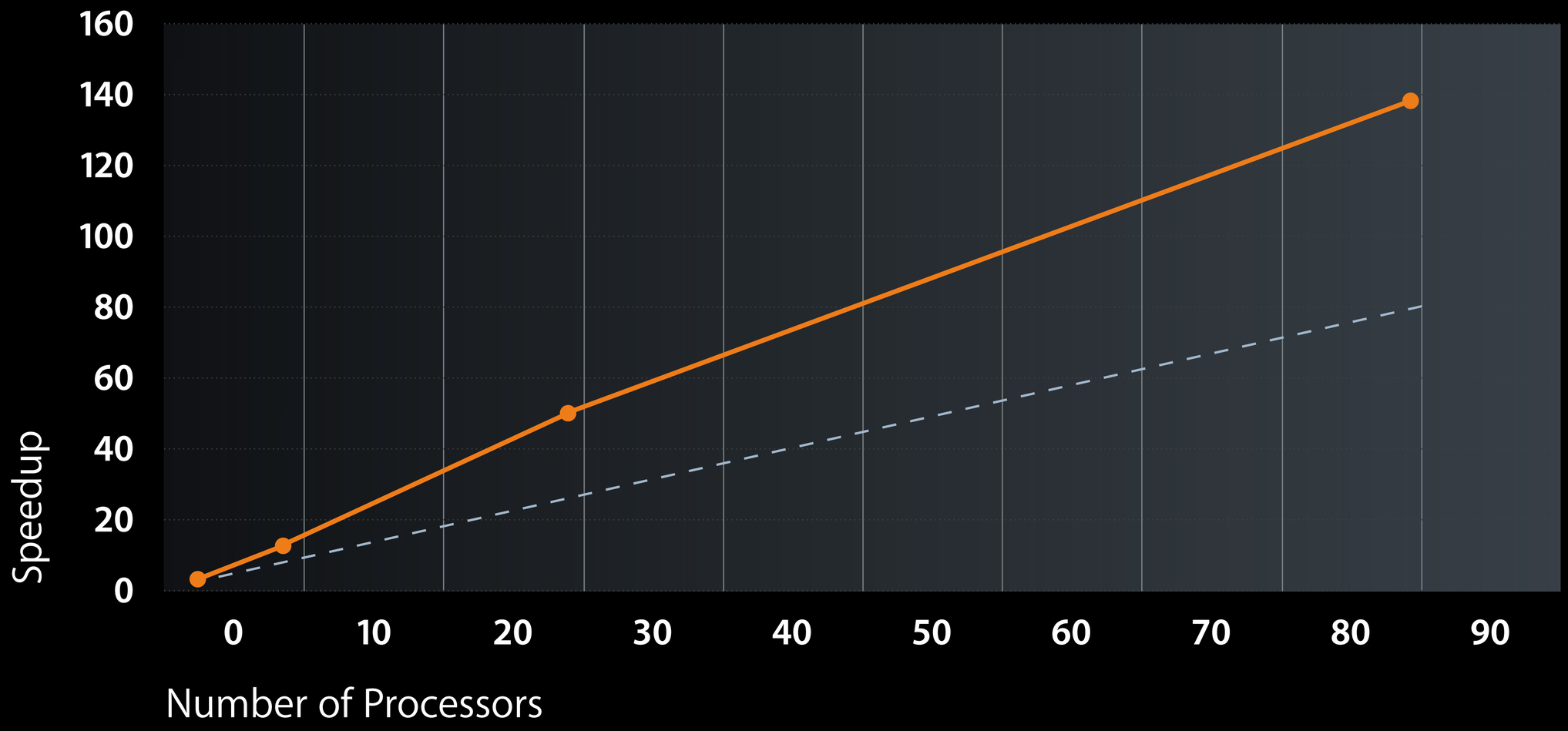
MPT HOWARD® Software Technology

Developer opportunities

- No need to rewrite applications—simply replace compute-intense functions with HOWARD® library functions
- Deliver higher parallelization performance using HOWARD® library functions
- Faster time to solution—HOWARD® library functions easy and intuitive to use

HOWARD® Speedup

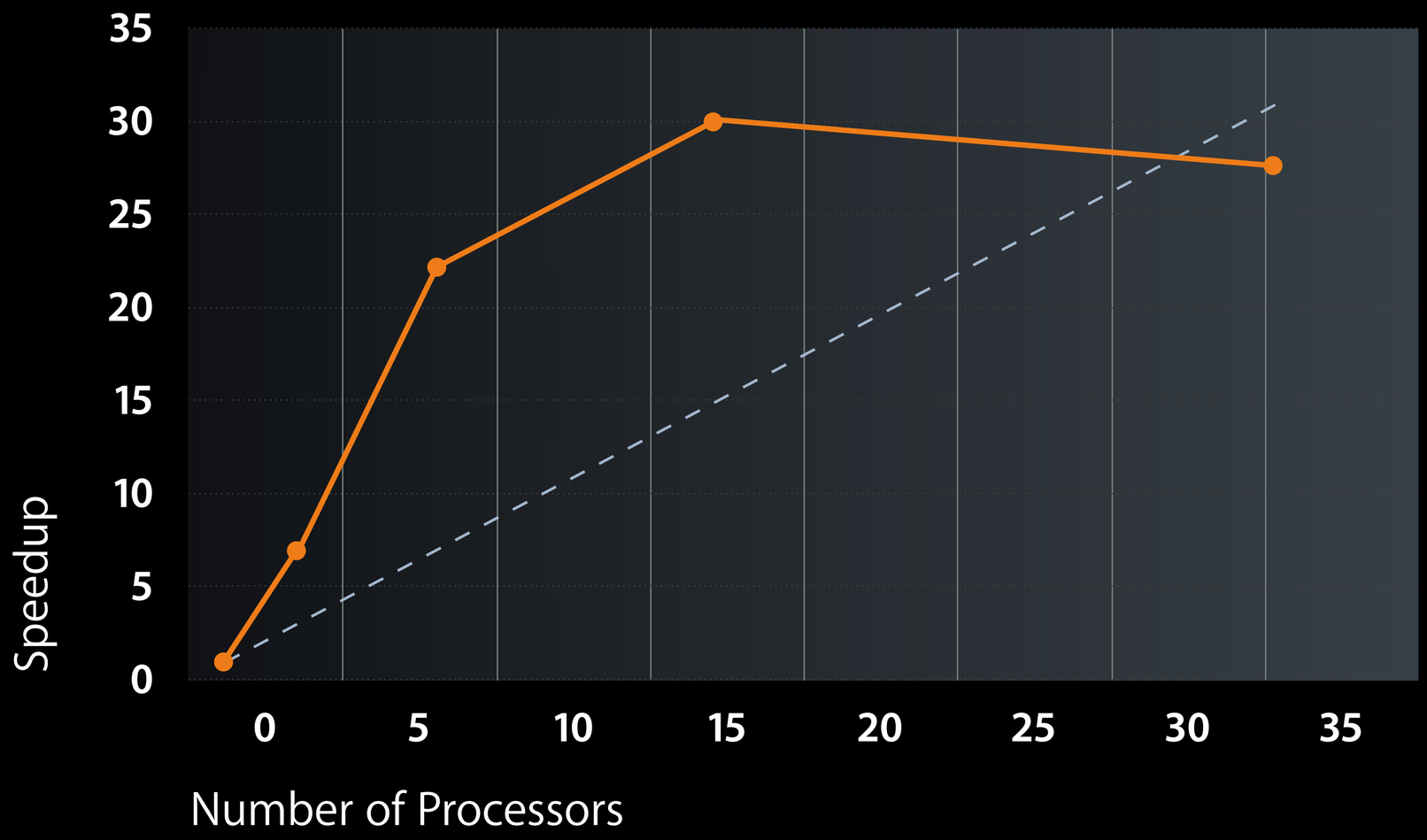
ex, 20,000 Digits of Accuracy, Two-Channels



U.S. Patent #6,857,004 Other Patents Pending

HOWARD® Speedup

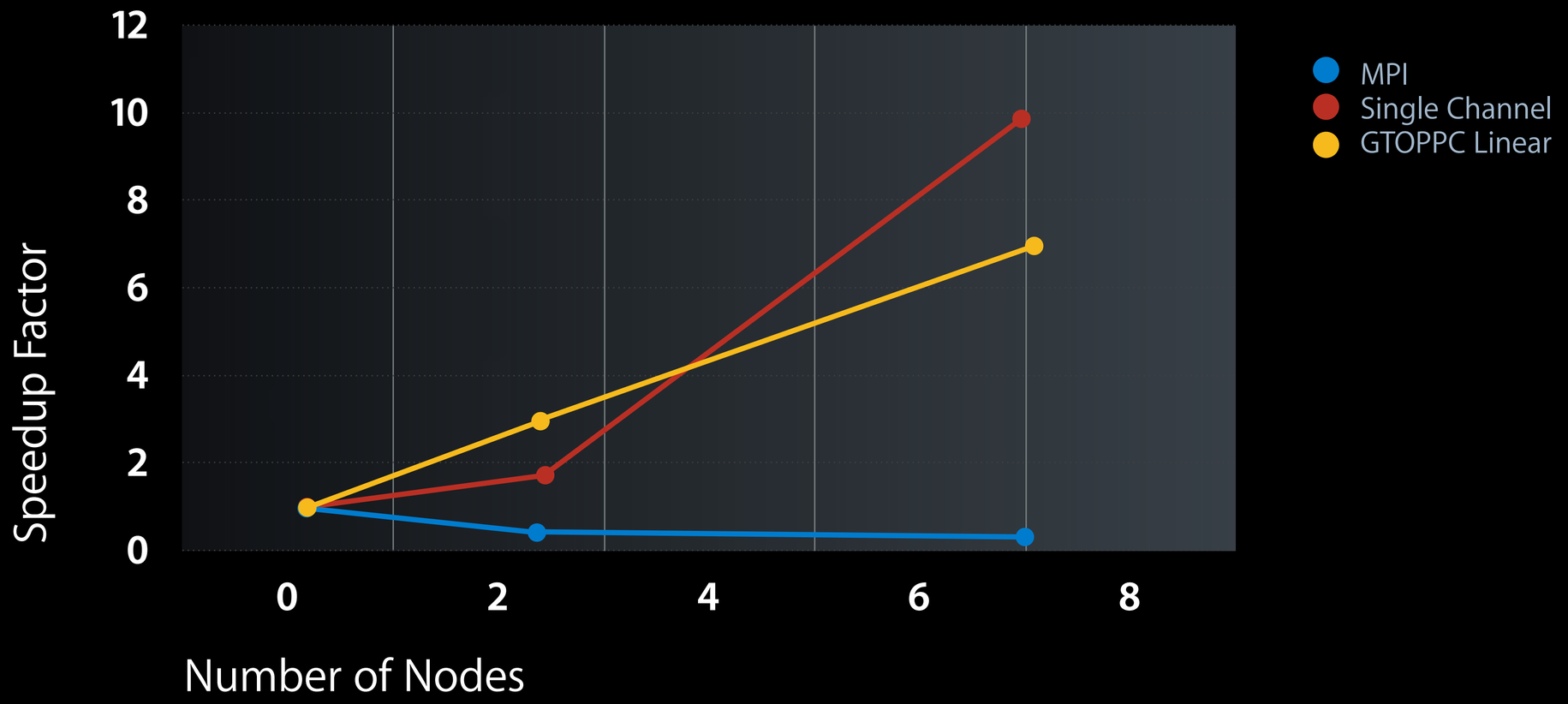
Triangulated Irregular Networks



U.S. Patent #6,857,004 Other Patents Pending

HOWARD® Speedup

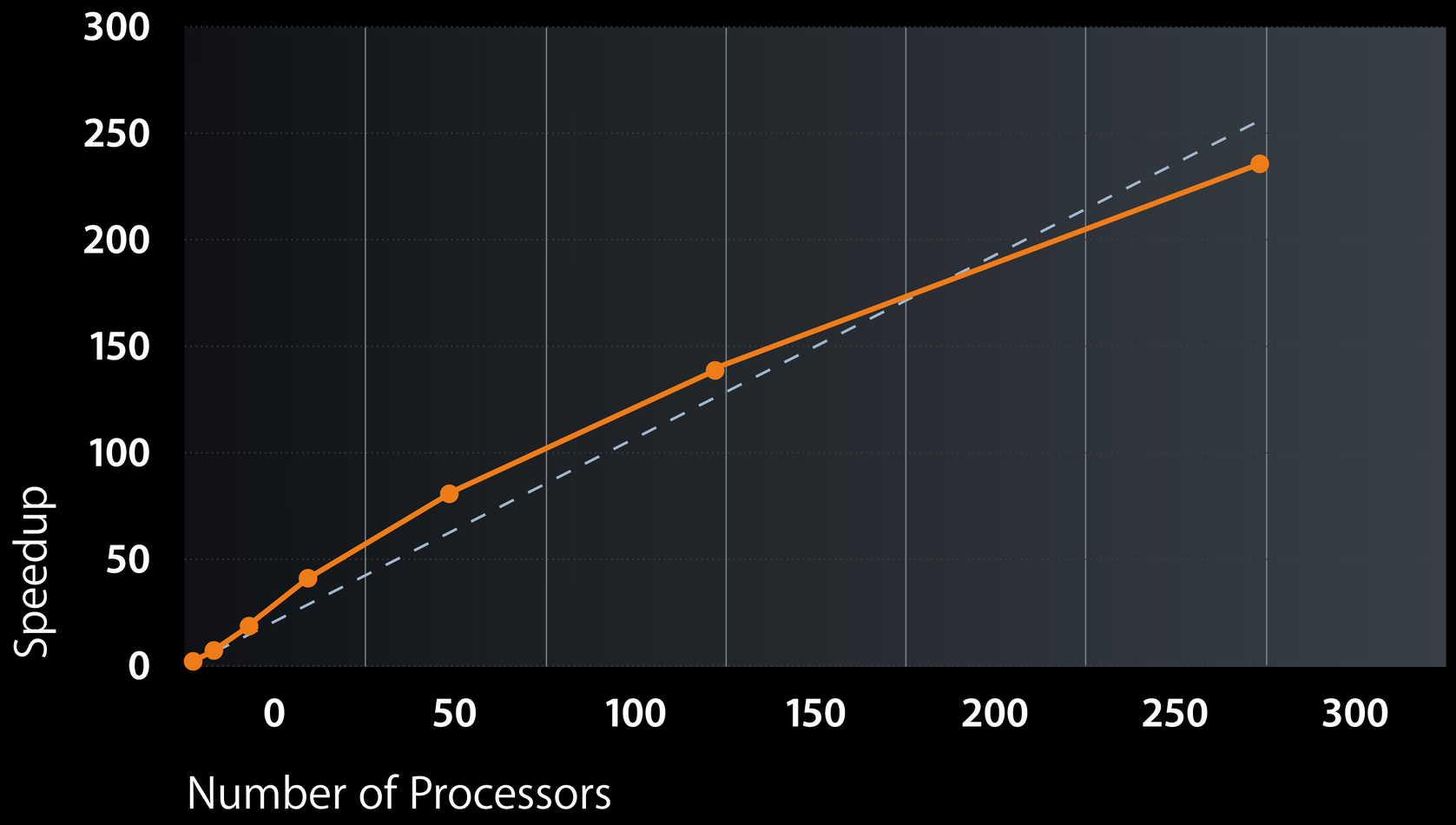
LU Decomposition, Speedup Comparison (Daisy)



U.S. Patent #6,857,004 Other Patents Pending

HOWARD® Speedup

NCBI-based blastp



U.S. Patent #6,857,004 Other Patents Pending

The Technology Breakthrough

Effective Bandwidth

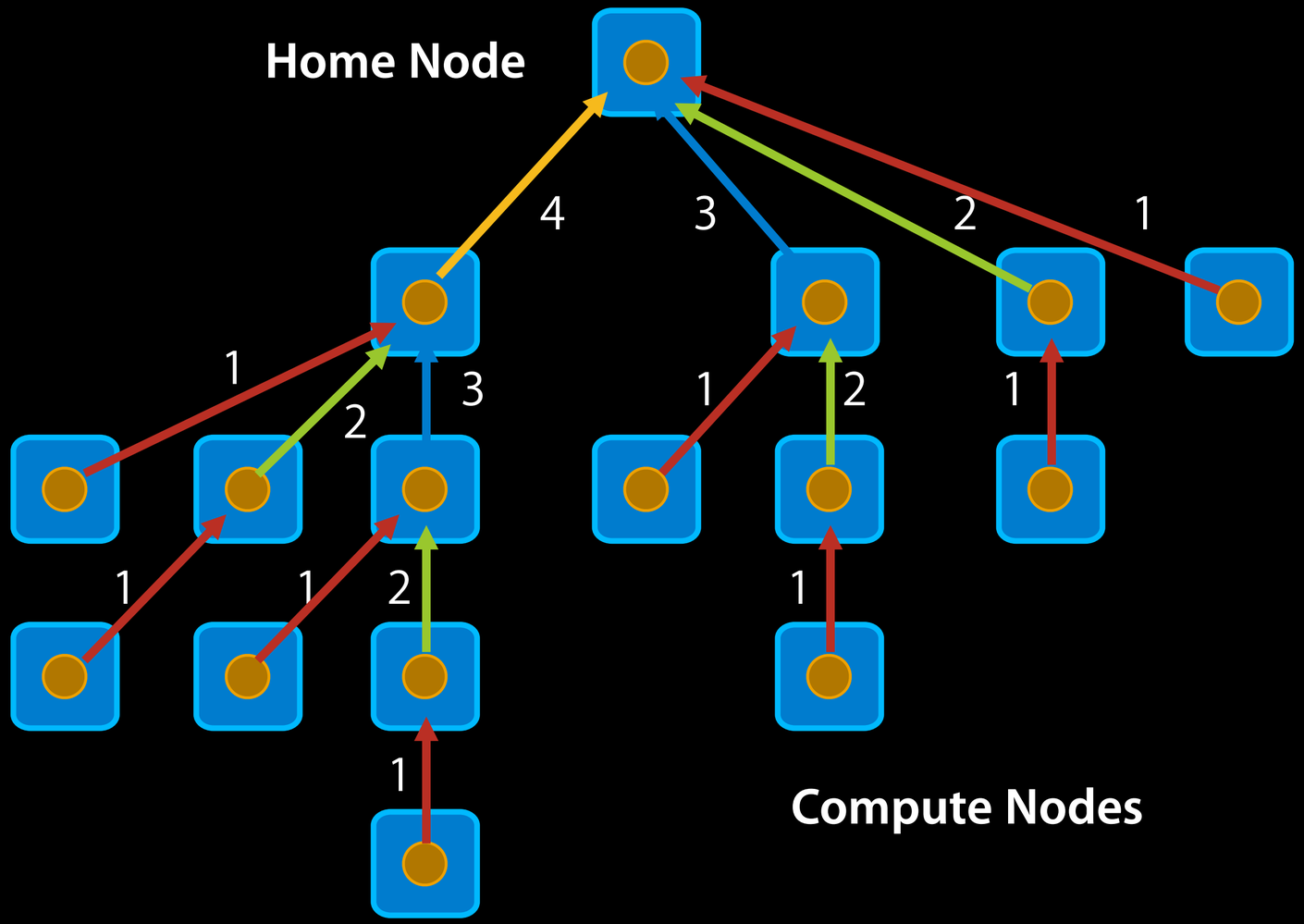
100 BaseT Point-to-Point Bandwidth

	Time Unit	Cascade A Effective Bandwidth	Cascade B Effective Bandwidth	Cascade C Effective Bandwidth	Manifold A Effective Bandwidth
1	100 Mb/s	200 Mb/s	200 Mb/s		
2	300 Mb/s	600 Mb/s	800 Mb/s	600 Mb/s	2.0 Gb/s
3	700 Mb/s	1.4 G/bs	2.6 G/bs	1.8 G/bs	8.0 G/bs
4	1.5 Gb/s	3.0 G/bs	8.0 G/bs	4.2 G/bs	26.0 G/bs
5	3.1 Gb/s	6.2 G/bs	24.2 G/bs	9.0 G/bs	80.0 G/bs
6	6.3 Gb/s	12.6 G/bs	72.8 G/bs	18.6 G/bs	242.0 G/bs
7	12.7 Gb/s	25.4 G/bs	218.6 G/bs	37.8 G/bs	2.2 G/bs
8	25.5 Gb/s	51.0 G/bs	656.0 G/bs	76.2 G/bs	6.6 G/bs

U.S. Patent #6,857,004 Other Patents Pending

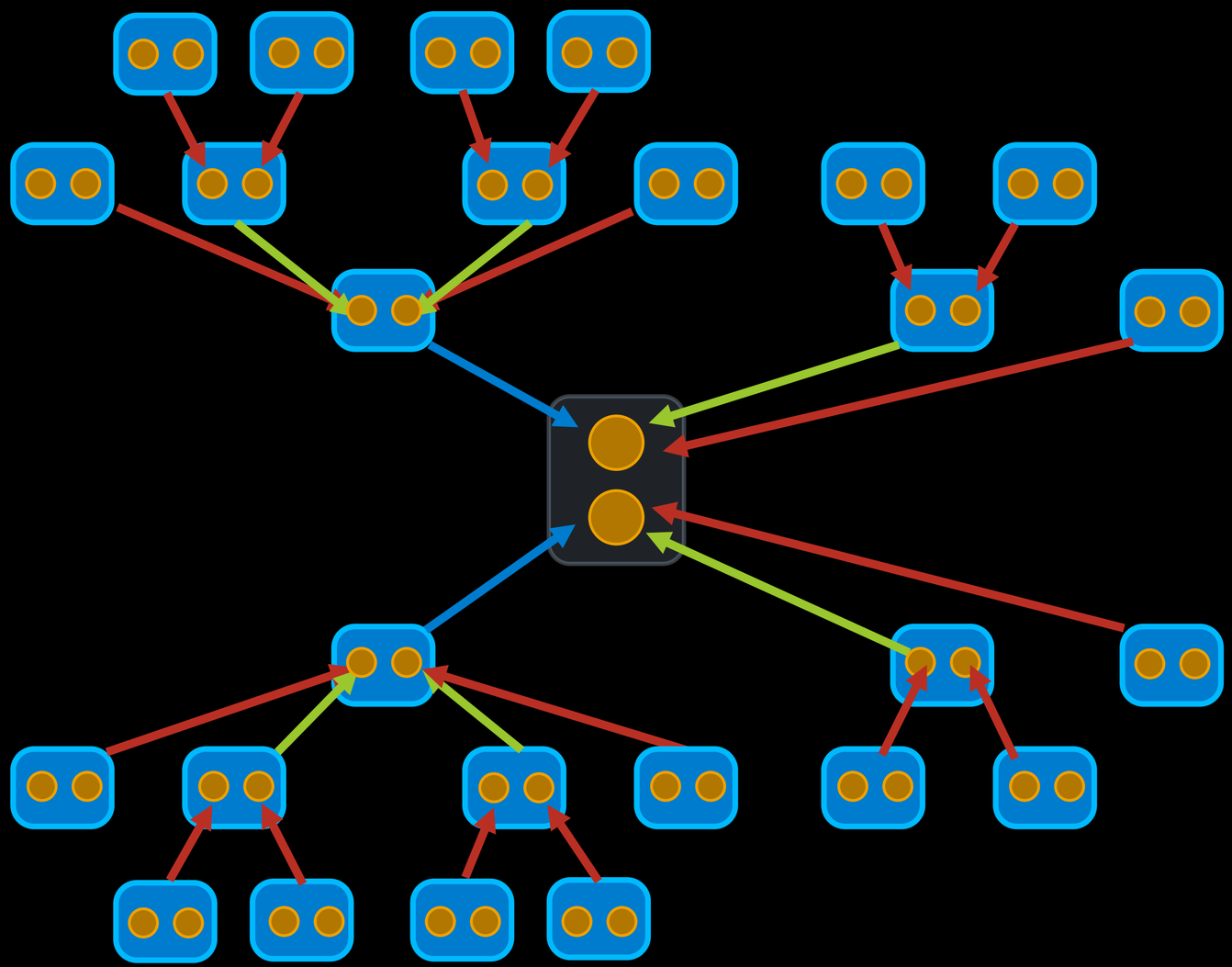
Technology Effects

HOWARD® Cascade, Single-Channel



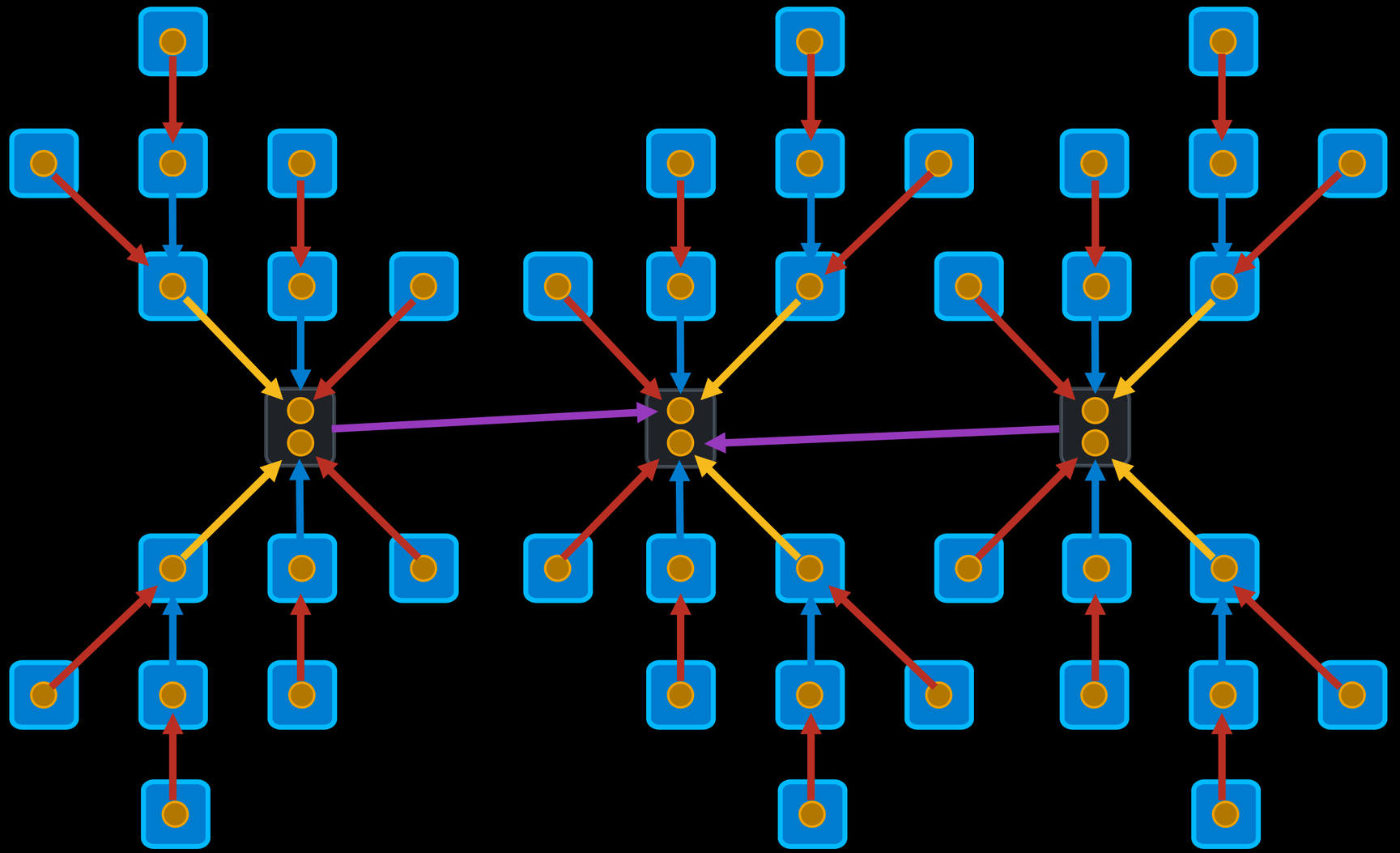
Technology Effects

HOWARD® Cascade, Two-Channel



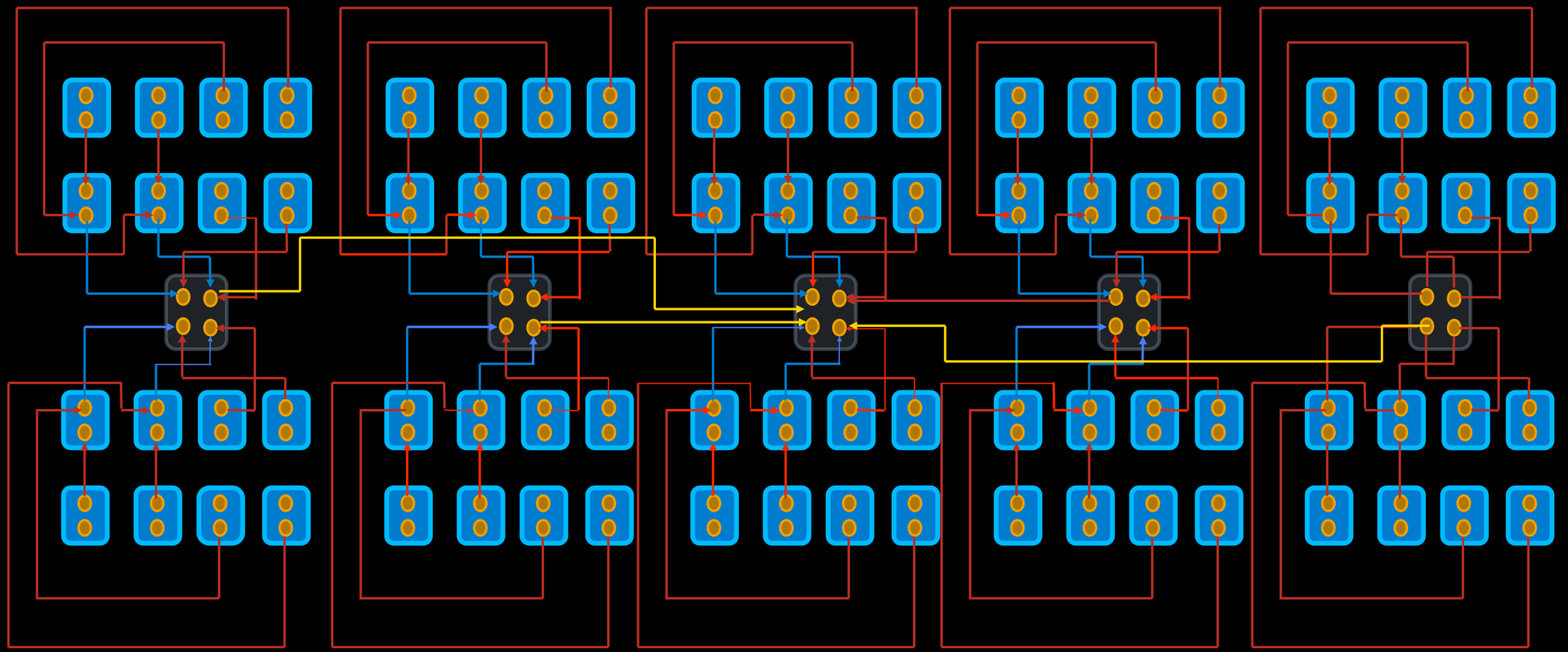
Technology Effects

HOWARD®-Lupo Manifold, Single-Channel, Four Time Units



Technology Effects

HOWARD®-Lupo Manifold, Two-Channel, Three Time Units



Massively Parallel Demo

http://developer.apple.com/wwdc2005/

MPT Applications and Markets

- Bioinformatics
- Molecular modeling
- Image processing
- Neural networks
- Seismic processing
- Computational fluid dynamics

Contact MPT

www.massivelyparallel.com
info@massivelyparallel.com
303-926-8555

More Information

http://developer.apple.com/wwdc2005/

Apple HPC Resources

- Performance and Optimization Workshops
- Apple System Engineers
- Developer Technical Support
- Driving an HPC application to Mac OS X
- Driving a technology to Mac OS X

Resources

Apple SciTech List
lists.apple.com/mailman/listinfo/scitech
scitech@lists.apple.com

Apple HPC List
lists.apple.com/mailman/listinfo/hpc
hpc@lists.apple.com

HPC page at developer.apple.com
http://developer.apple.com/hardware/hpc/

Apple in Science
apple.com/science

Related Sessions

Large Installation and HPC Systems Administration	Session 618 Wednesday 3:30PM
Using HPC: Infiniband Cluster	Lab 6008 Wednesday 5:00PM
Using Xgrid to Create and Deploy Distributed Computations	Session 616 Wednesday 2:00PM
Using Xgrid	Lab 6018 Friday 2:00PM
Debugging Parallel Applications	Session 643 Friday 5:00PM

Who to Contact

Skip Levens
Server Technology Manager
skip@apple.com

WWDC2005

Worldwide Developers Conference